

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-235400

(43)Date of publication of application : 29.08.2000

(51)Int.Cl.

G10L 19/12

G10L 13/00

H04B 14/04

(21)Application number : 11-035420

(71)Applicant : NIPPON TELEGR & TELEPH CORP  
<NTT>

(22)Date of filing : 15.02.1999

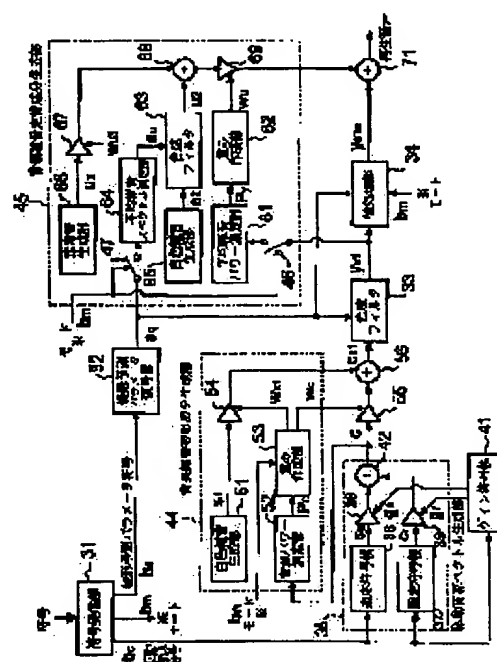
(72)Inventor : OMURO NAKA  
MANO KAZUNORI

(54) ACOUSTIC SIGNAL CODING DEVICE, DECODING DEVICE, METHOD FOR THESE AND PROGRAM RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To realize a natural reproducing sound without reproducing an unpleasant sound by sending a characteristic of a background noise caught as a short time characteristic and a long time characteristic on a transmission parameter and reproducing both characteristics in a reproducing side.

SOLUTION: A reproduced sound signal is constituted of a whole or a part of signals generated in a drive sound source vector generation part 35, a background noise fluctuation component generation part 44 and a background noise steady component generation part 45. A model of a noise added to a regenerative signal expressed by a generation model of a sound is expressed with two components of a 'background noise fluctuation component' that a characteristic is fluctuated in a short time, and a 'background noise steady component' that the characteristic is fluctuated at a hourly slow period, or isn't fluctuated. A vector outputted from the background noise fluctuation component generation part 44 answers to the component that power and a spectrum characteristic are fluctuated by the unit of the frame of a background noise. The level of the background noise steady component is decided by a measured result of mean noise power, and is added with a nearly fixed level regardless of a mode at every frame.



## LEGAL STATUS

[Date of request for examination]

23.05.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision  
of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japanese Patent Office

出 願 特願平11-35420 (平11. 2.15)

公 開 特開2000-235400 (平12. 8.29)

公 告

登 録

名 称 音響信号符号化装置、復号化装置、これらの方法、及びプログラム記録媒体

抄 録 【要約】 【課題】 CELP方式で雑音環境の符号化でも自然な再生音を得る。  
 【解決手段】 当該フレームが背景雑音区間であるか否かの情報を受信し、背景雑音区間で、合成フィルタ33の出力のパワーの長時間平均値を表わすパワーレベルを求め(61)、また復号線形予測パラメータからその長時間平均を表わす平均スペクトルを求め(64)、この平均スペクトルをフィルタ係数とする合成フィルタ63を白色雑音で駆動し、そのフィルタ出力をパワーレベルを振幅調整して背景雑音の定常成分信号を得、この信号を合成フィルタ33の出力に加算する。

出願人 日本電信電話株式会社

発明者 大室仲  
間野一則

IPC G10L 19/12

G10L 13/00

H04B 14/04

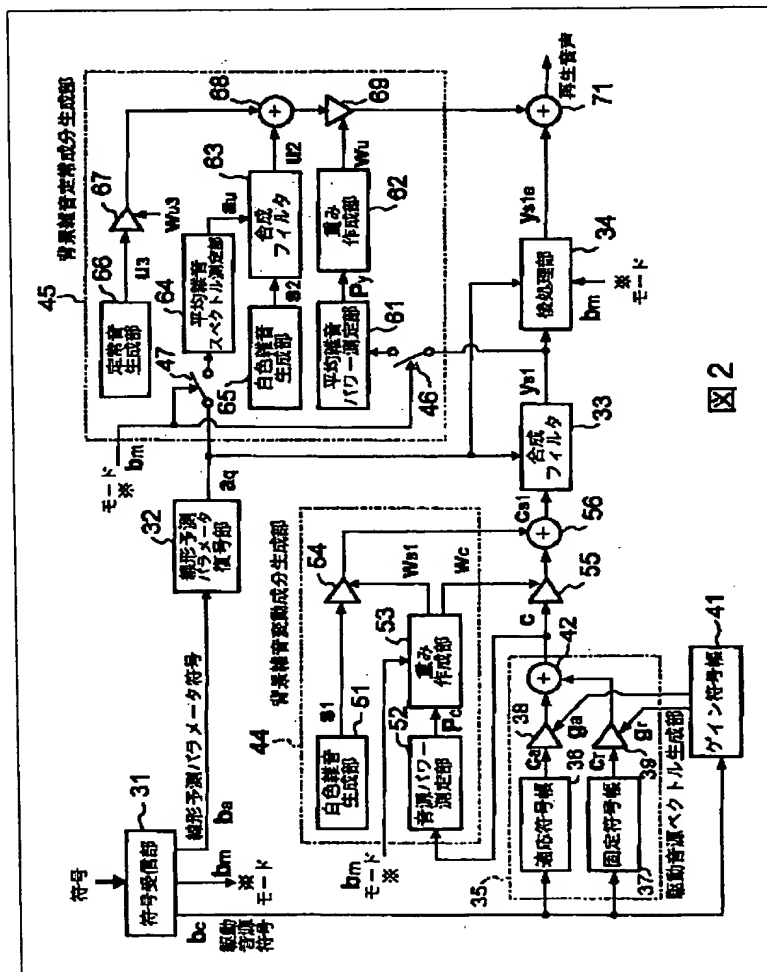


図 2

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2000-235400  
(P2000-235400A)

(43) 公開日 平成12年8月29日 (2000.8.29)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テマコード* (参考)
G 1 0 L 19/12		G 1 0 L 9/14	S 5 D 0 4 5
13/00			L 5 K 0 4 1
H 0 4 B 14/04		H 0 4 B 14/04	Z 9 A 0 0 1

審査請求 未請求 請求項の数15 O L (全 13 頁)

(21) 出願番号 特願平11-35420

(22) 出願日 平成11年2月15日 (1999.2.15)

(71) 出願人 000004226

日本電信電話株式会社  
東京都千代田区大手町二丁目3番1号

(72) 発明者 大室 伸

東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(72) 発明者 間野 一則

東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(74) 代理人 100066153

弁理士 草野 卓 (外1名)

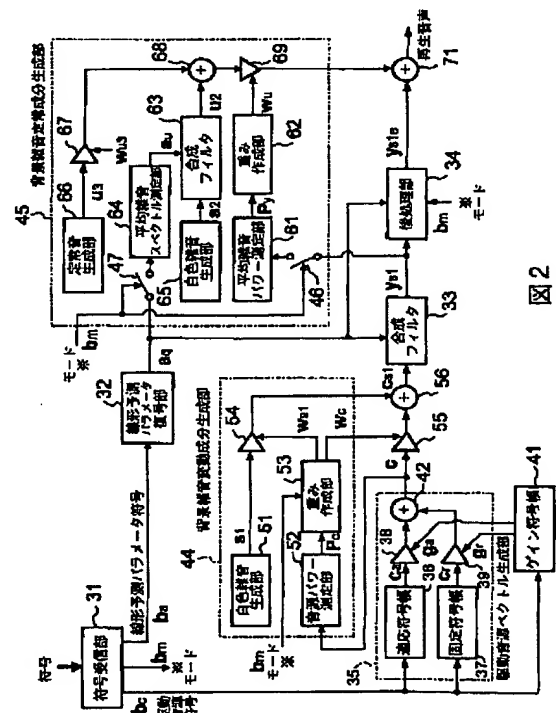
最終頁に続く

(54) 【発明の名称】 音響信号符号化装置、復号化装置、これらの方法、及びプログラム記録媒体

(57) 【要約】

【課題】 CELP方式で雑音環境の符号化でも自然な再生音を得る。

【解決手段】 当該フレームが背景雑音区間であるか否かの情報を受信し、背景雑音区間で、合成フィルタ33の出力のパワーの長時間平均値を表わすパワーレベルを求め(61)、また復号線形予測パラメータからその長時間平均を表わす平均スペクトルを求め(64)、この平均スペクトルをフィルタ係数とする合成フィルタ63を白色雑音で駆動し、そのフィルタ出力をパワーレベルを振幅調整して背景雑音の定常成分信号を得、この信号を合成フィルタ33の出力に加算する。



## 【特許請求の範囲】

【請求項1】 適応符号帳と固定符号帳の少くとも一方からフレーム単位で符号ベクトルを選択し、ゲイン符号帳から選択されたゲインを前記選択された符号ベクトルに乗算して駆動音源ベクトルを生成し、前記駆動音源ベクトルで合成フィルタを駆動して音響信号を出力する音響信号復号化装置において、当該フレームが背景雑音区間であるか否かを示す情報を受信する手段と、背景雑音区間であるフレームにおいて前記合成フィルタからの出力信号パワーの長時間平均値を表すパワーレベルを計算する手段と、背景雑音区間であるフレームにおいて前記合成フィルタのフィルタ係数よりスペクトルの長時間平均を表す平均スペクトルを計算する手段と、白色雑音で駆動され、前記平均スペクトルがフィルタ係数として与えられる第2合成フィルタと、前記第2合成フィルタで生成された信号を、前記計算されたパワーレベルに基づいて振幅調整して背景雑音の定常成分信号を生成する手段と、前記定常成分信号を、当該フレームが背景雑音区間であるか否かに関わらず前記合成フィルタからの出力信号に加算して再生音響信号を出力する手段と、を有する音響信号復号化装置。

【請求項2】 一定のスペクトル特性及び一定の振幅を有する定常音を生成する手段と、前記背景雑音の定常成分信号を生成する手段よりの定常成分信号と前記定常音を加算して定常成分信号として出力する手段とを備えることを特徴とする請求項1記載の音響信号復号化装置。

【請求項3】 当該フレームが背景雑音区間又は子音区間であるか否かを示す情報を受信する手段と、前記駆動音源ベクトルのパワーに基づいて重みを決定する手段と、前記重みを白色雑音に乗算して背景雑音の変動成分信号を生成する手段と、背景雑音区間又は子音区間であるフレームにおいて、前記変動成分信号を前記駆動音源ベクトルに加算する手段と、を有することを特徴とする請求項1又は2記載の音響信号復号化装置。

【請求項4】 適応符号帳と固定符号帳の少くとも一方からフレーム単位で符号ベクトルを選択し、ゲイン符号帳から選択されたゲインを前記選択された符号ベクトルに乗算して駆動音源ベクトルを生成し、前記駆動音源ベクトルで合成フィルタを駆動して得られた音響信号と入力音響信号とを比較して前記符号ベクトルと前記ゲインを選択する音響信号符号化装置において、入力音響信号を分析して、当該フレームが背景雑音区間

又は子音区間に相当するか否かを決定する手段と、前記合成フィルタの出力信号と入力音響信号とのパワーレベル差を計算する手段と、前記ゲイン選択に際し、背景雑音区間又は子音区間のフレームにおいて前記パワーレベル差を距離尺度とし、その距離尺度を最小とするゲインをゲイン符号帳から選択する手段と、を備えることを特徴とする音響信号符号化装置。

【請求項5】 前記距離尺度として、前記合成フィルタからの出力信号と入力音響信号との波形歪みと、前記パワーレベル差との加重和が用いられることを特徴とする請求項4記載の音響信号符号化装置。

【請求項6】 適応符号帳と固定符号帳の少くとも一方からフレーム単位で符号ベクトルを選択し、ゲイン符号帳から選択されたゲインを前記選択された符号ベクトルに乗算して駆動音源ベクトルを生成し、前記駆動音源ベクトルで合成フィルタを駆動して音響信号を出力する音響信号復号化方法において、当該フレームが背景雑音区間であるか否かを示す情報を受信する過程と、背景雑音区間であるフレームで前記合成フィルタからの出力信号パワーの長時間平均値を表すパワーレベルを計算する過程と、前記合成フィルタのフィルタ係数よりスペクトルの長時間平均を表す平均スペクトルを計算する過程と、前記平均スペクトルをフィルタ係数とする第2合成フィルタを白色雑音で駆動する過程と、前記第2合成フィルタで生成された信号を、前記計算されたパワーレベルに基づいて振幅調整して背景雑音の定常成分信号を生成する過程と、前記定常成分信号を、当該フレームが背景雑音区間であるか否かに関わらず前記合成フィルタからの出力信号に加算して音響信号を出力する過程と、を有する音響信号復号化方法。

【請求項7】 一定のスペクトル特性及び一定の振幅を有する定常音を生成する過程と、前記背景雑音の定常成分信号と前記定常音を加算して定常成分信号として出力する過程とを有する、ことを特徴とする請求項6記載の音響信号復号化方法。

【請求項8】 当該フレームが背景雑音区間又は子音区間であるか否かを示す情報を受信する過程と、前記駆動音源ベクトルのパワーに基づいて重みを決定する過程と、前記重みを白色雑音に乗算して背景雑音の変動成分信号を生成する過程と、背景雑音区間又は子音区間のフレームで、前記変動成分信号を前記駆動音源ベクトルに加算する過程と、を有することを特徴とする請求項6又は7記載の音響信号復号化方法。

【請求項9】 適応符号帳と固定符号帳の少くとも一方

からフレーム単位で符号ベクトルを選択し、各々ゲイン符号帳から選択されたゲインを前記選択された符号ベクトルに乘算して駆動音源ベクトルを生成し、前記駆動音源ベクトルで合成フィルタを駆動して得られた音響信号と入力音響信号とを比較して前記符号ベクトルと前記ゲインを選択する音響信号符号化方法において、

入力音響信号を分析して、当該フレームが背景雑音区間又は子音区間に相当するか否かを決定する過程と、前記合成フィルタの出力信号と入力音響信号とのパワーレベル差を計算する過程と、前記ゲイン選択に際し、背景雑音区間又は子音区間のフレームで前記パワーレベル差を距離尺度とし、該距離尺度を最小とするゲインをゲイン符号帳から選択する過程と、を有することを特徴とする音響信号符号化方法。

【請求項10】 前記距離尺度として、前記合成フィルタからの出力信号と入力音響信号との波形歪みと、前記パワーレベル差との加重和を用いることを特徴とする請求項9記載の音響信号符号化方法。

【請求項11】 適応符号帳と固定符号帳の少くとも一方からフレーム単位で符号ベクトルを選択し、各々ゲイン符号帳から選択されたゲインを前記選択された符号ベクトルに乘算して駆動音源ベクトルを生成し、前記駆動音源ベクトルで合成フィルタを駆動して音響信号を出力する音響信号復号化装置のコンピュータに、当該フレームが背景雑音区間であるか否かを示す情報を受信する処理と、背景雑音区間のフレームで前記合成フィルタからの出力信号パワーの長時間平均値を表すパワーレベルを計算する処理と、前記合成フィルタのフィルタ係数よりスペクトルの長時間平均を表す平均スペクトルを計算する処理と、白色雑音で前記平均スペクトルをフィルタ係数とする第2合成フィルタを白色雑音で駆動する処理と、前記第2合成フィルタで生成された信号を、前記計算されたパワーレベルに基づいて振幅調整して背景雑音の定常成分信号を生成する処理と、前記定常成分信号を、当該フレームが背景雑音区間であるか否かに関わらず前記合成フィルタからの出力信号に加算して音響信号を出力する処理と、を実行させるプログラムを記憶した記録媒体。

【請求項12】 一定のスペクトル特性及び一定の振幅を有する定常音を生成する処理と、前記背景雑音の定常成分信号と前記定常音を加算して定常成分信号として出力する処理とを、前記コンピュータに実行させるプログラムを前記プログラムを含むことを特徴とする請求項11記載の記録媒体。

【請求項13】 当該フレームが背景雑音区間又は子音

区間であるか否かを示す情報を受信する処理と、前記駆動音源ベクトルのパワーに基づいて重みを決定する処理と、前記重みを白色雑音に乘算して背景雑音の変動成分信号を生成する処理と、背景雑音区間又は子音区間のフレームで、前記変動成分信号を前記駆動音源ベクトルに加算する処理と、を前記コンピュータに実行させるプログラムを前記プログラムを含むことを特徴とする請求項11又は12記載の記録媒体。

【請求項14】 適応符号帳と固定符号帳の少くとも一方からフレーム単位で符号ベクトルを選択し、各々ゲイン符号帳から選択されたゲインを前記選択された符号ベクトルに乘算して駆動音源ベクトルを生成し、前記駆動音源ベクトルで合成フィルタを駆動して得られた音響信号と入力音響信号とを比較して前記符号ベクトルと前記ゲインを選択する音響信号符号化装置のコンピュータに、入力音響信号を分析して、当該フレームが背景雑音区間又は子音区間に相当するか否かを決定する処理と、前記合成フィルタの出力信号と入力音響信号とのパワーレベル差を計算する処理と、前記ゲイン選択に際し、背景雑音区間又は子音区間のフレームで前記パワーレベル差を距離尺度とし、該距離尺度を最小とするゲインをゲイン符号帳から選択する処理と、を実行するプログラムを記憶した記録媒体。

【請求項15】 前記距離尺度として、前記合成フィルタからの出力信号と入力音声信号との波形歪みと、前記パワーレベル差との加重和を求める処理を前記コンピュータに実行させるプログラムを前記プログラムに含むことを特徴とする請求項14記載の記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、音声をはじめとする音響信号の信号系列を、少ない情報量でデジタル符号化する高能率音響信号符号化方法、その復号化方法、これらの装置に関する。

【0002】

【従来の技術】デジタル移動体通信において、電波を効率的に利用したり、音声または音楽蓄積サービス等で通信回線や記憶媒体を効率的に利用するために、高能率音声符号化方法が用いられる。現在、音声を高能率に符号化する方法として、原音声を、フレームまたはサブフレーム（以降総称してフレーム）と呼ばれる5～50ms程度の一定間隔の区間に分割し、その1フレームの音声を、周波数スペクトルの包絡特性を表す線形フィルタの特性と、そのフィルタを駆動するための駆動音源信号との2つの情報に分離し、それぞれを符号化する手法が提案されている。この手法において、駆動音源信号を符

号化する方法として、音声のピッチ周期（基本周波数）に対応すると考えられる周期成分と、それ以外の成分に分離して符号化する方法が知られている。この駆動音源情報の符号化法の例として、符号駆動線形予測符号化（Code-Excited Linear Prediction: CELP）がある。この技術の詳細については、文献M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rate", IEEE Proc. ICASSP-85, pp.937-940, 1985に記載されている。

【0003】図3に上記符号化方法の構成例を示す。入力端子11に入力された音声 $x$ は、線形予測分析部12において、入力音声の周波数スペクトル包絡特性を表す線形予測パラメータ $a$ が計算される。得られた線形予測パラメータ $a$ は線形予測パラメータ符号化部13において、量子化および符号化され、量子化値は合成フィルタ係数 $a_q$ に変換されて合成フィルタ14に、符号 $b_a$ は符号送出部15へ送られる。なお、歪み計算に聴覚特性を考慮するなど、入力音声のスペクトル情報を利用して歪み計算を行う場合には、線形予測パラメータ $a$ または量子化された線形予測パラメータ $a_q$ が波形歪み計算部16へも送られる。線形予測分析の詳細および線形予測パラメータの符号化例については、例えば古井貞熙著“デジタル音声処理”（東海大学出版会）に記載されている。ここで、線形予測分析部12、線形予測パラメータ符号化部13および合成フィルタ14は非線形なものに置き換えてもよい。

【0004】駆動音源ベクトル生成部17では、1フレーム分の長さの駆動音源ベクトル候補を生成し、合成フィルタ14に送る。駆動音源ベクトル生成部17は、一般に適応符号帳18と固定符号帳19から構成されることが多い。適応符号帳18からはバッファに記憶された直前の過去の駆動音源ベクトル（既に量子化された直前の1～数フレーム分の駆動音源ベクトル）を、ある周期に相当する長さで切り出し、その切り出したベクトルをフレームの長さになるまで繰り返すことによって、音声の周期成分に対応する時系列ベクトルの候補が出力される。上記「ある周期」とは、波形歪み計算部16における歪みが小さくなるような周期が選択されるが、選択された周期は、一般には音声のピッチ周期に相当することが多い。固定符号帳19からは、音声の非周期成分に対応する1フレーム分の長さの時系列符号ベクトルの候補が出力される。これらの候補は入力音声とは独立に符号化のためのビット数に応じて、あらかじめ指定された数の候補ベクトルを記憶してそのうちの1つであったり、あらかじめ決められた生成規則によってパルスを配置して生成されたベクトルの1つであったりする。なお、固定符号帳19は、本来音声の非周期成分に対応するものであるが、特に母音区間など、ピッチ周期性の強い音声区間では、上記あらかじめ用意された候補ベクトルに、

ピッチ周期または適応符号帳で用いるピッチに対応する周期を持つ櫛形フィルタをかけたり、適応符号帳での処理と同様にベクトルを切り出して繰り返したりして固定符号ベクトルとすることもある。適応符号帳18および固定符号帳19から出力された時系列ベクトルの候補

$c_a$ および $c_r$ は、乗算部21および22において、それぞれゲイン符号帳23から出力されるゲイン候補 $g_a$ 、 $g_r$ が乗算され、加算部24において加算され、駆動音源ベクトルの候補 $c$ となる。図3の構成例において、実際の動作中には適応符号帳18のみまたは固定符号帳19のみが用いられる場合もある。

【0005】合成フィルタ14は、線形予測パラメータ符号化部13において量子化された線形予測パラメータ $a_q$ から得られる合成フィルタ係数をフィルタの係数とする線形フィルタであって、駆動音源ベクトル候補 $c$ を入力として再生音声の候補 $y$ を出力する。合成フィルタ14の次数すなわち線形予測分析の次数は、一般に10～16次程度が用いられることが多い。なお、既に述べたように、合成フィルタ14は非線形なフィルタでもよい。

【0006】波形歪み計算部16では、合成フィルタ14の出力である再生音声の候補 $y$ と、入力音声 $x$ との歪み $d$ を計算する。この歪みの計算は、例えば聴覚重み付けに代表されるように、合成フィルタ14の係数 $a_q$ または量子化していない線形予測係数 $a$ を考慮にいれて行なうことが多い。符号帳検索制御部25では、各再生音声候補 $y$ と入力音声 $x$ との歪みが最小または最小に準ずるような駆動音源符号 $b_c$ 、すなわち周期符号、固定（雑音）符号およびゲイン符号を選択し、そのフレームにおける駆動音源ベクトルを決定する。

【0007】符号帳検索制御部25において決定された駆動音源符号 $b_c$ （周期符号、固定符号、ゲイン符号）と、線形予測パラメータ符号化部13の出力である線形予測パラメータ符号 $b_a$ は、符号送出部15に送られ、利用の形態に応じて記憶装置に記憶されるか、または通信路を介して受信側へ送られる。図4に、上記符号化方法に対応する復号方法の構成例を示した。伝送路または記憶媒体から符号受信部31において受信された符号のうち、線形予測パラメータ符号 $b_a$ は線形予測パラメータ復号部32において合成フィルタ係数 $a_q$ に復号され、合成フィルタ33および、必要に応じて後処理部（ポストフィルタとも呼ばれる）34に送られる。受信された符号のうち、駆動音源符号 $b_c$ は、駆動音源ベクトル生成部35に送られ、符号に対応する音源ベクトル $c$ が生成される。合成フィルタ33は、駆動音源ベクトル $c$ を入力として、合成音声 $y$ を出力し、後処理部34はスペクトル強調やピッチ強調の処理を合成音声 $y$ に施して、量子化ノイズを聴覚的に低減する。なお、後処理部34は一種の音声強調処理である

ので、処理量の関係や、入力信号の特性によって用いない場合もある。駆動音源ベクトル生成部35は駆動音源符号  $\mathbf{b}$  c 中の周期符号により適応符号帳36から時系列ベクトル  $\mathbf{c}$  a が選択され、また固定符号により固定符号帳37から時系列ベクトル  $\mathbf{c}$  r が選択され、これら時系列ベクトル  $\mathbf{c}$  a,  $\mathbf{c}$  r は乗算部38, 39で、ゲイン符号によりゲイン符号帳41から取り出されたゲイン  $\mathbf{g}$  a,  $\mathbf{g}$  r が乗算されて加算部42で互いに加算されて駆動音源ベクトル  $\mathbf{c}$  として合成フィルタ33に入力される。前述のように実際の動作中に、適応符号帳18のみ又は固定符号帳19のみが符号化に用いられる場合には、それに対応して、図4においては適応符号帳36又は固定符号帳37のみが用いられる。

#### 【0008】

【発明が解決しようとする課題】このようなCELP系符号化方式をはじめとする音声の生成モデルに基づく符号化方式において問題となるのは、静かな環境で録音された背景雑音のない音声信号が入力された場合には、少ない情報量で高品質な符号化を実現することができるが、オフィスや街頭など、背景雑音のある環境で録音された音声が入力されると、キュルキュルとかバチバチといった大変に不快な音が再生されるといった点である。これらの背景雑音を入力した場合の問題は、ピッチ周期性を利用するCELP系の音声符号化モデルが音声の生成モデルに基づいているのに対して、背景雑音は音声とは異なる性質を示すためである。具体的には、適応符号帳が音声のピッチ周期に対応する信号成分を出力するのに対して、背景雑音には一般にピッチ周期性がないため、背景雑音区間において、不自然な周期音が発生する。また、背景雑音が重畳された音声区間においては、本来はピッチ周期性のある音声とピッチ周期性のない雑音信号が加算された性質の信号であるにもかかわらず、音声のピッチ周期性を重視する符号化モデルを適用するために、やはり背景雑音成分が不自然な周期音となって重畳する。固定符号帳をピッチ周期で周期化して用いる場合には、固定符号帳のピッチ周期化もまた不自然な周期音の発生する原因となる。上記のように、適応符号帳や固定符号帳の構成が信号の性質に合わない場合には、ゲインの決定方法にも問題が生じる。つまり、従来のゲインの決定方法は、適応符号帳や固定符号帳から出力される駆動音源ベクトルの性質が、入力信号の性質によく合っていることを前提としたものであり、駆動音源ベクトルの性質が入力信号の性質に合わない場合に、従来の方法では不自然に変動する信号となる。

【0009】この問題に対する代表的な解決法として、ノイズリダクションによる方法と、コンフォートノイズジェネレータという方法がある。前者は、入力信号の前に、雑音低減処理を入れて、背景雑音成分を相対的に低減するもので、雑音成分が低減された分だけ再生音における不快音も低減される。しかしながら、雑音低減処理

を入れても、完全に雑音がなくなるわけではなく、不快音をなくしてしまうことはできない。また、背景雑音が非定常音の場合には、十分な雑音低減効果そのものを得ることが難しい。一方、後者のコンフォートノイズジェネレータは、音声区間についてはCELP系符号化方式でそのまま符号化し、雑音区間については、適当な「心地よい」雑音、例えば白色雑音などを生成して置き換えるというものである。コンフォートノイズジェネレータの方法を使うと、キュルキュルといった不快な音は再生されなくなるが、オフィスや雑踏などさまざまな背景雑音の性質に対して、再生される音の雑音区間は、常に同じ性質の雑音になってしまっていて、背景音の情報は受信側に伝達されないという問題が生じる。また、背景雑音のレベルが高いときには、音声区間と雑音区間を誤りなく切り替えることは難しく、区間検出誤りが原因で逆に再生音が劣化してしまう場合や、区間検出誤りがなくても音声区間と雑音区間の性質に違いがありすぎて不連続な感じに聞こえてしまうことも多かった。

【0010】この発明では、CELP系の方式をはじめとする、音声の生成モデルに基づく音声符号化方式において、不快な音が再生されないでかつ、背景音の性質を受信側に伝えて、より自然な再生音を実現する符号化および復号する方法及びその装置を提供することにある。

#### 【0011】

【課題を解決するための手段】この発明では、背景雑音の特性を、例えばガタンという音や車が通過した音、足音、遠くでの人の話し声などの短時間特性（短時間変動成分）と、例えば定常的にざわざわした感じとかモーターの回転音などの平均的な長時間特性（長時間変動成分）という2つの立場でとらえ、両特性をCELP系符号化モデルの枠組みの中で送信パラメータに情報を乗せて受信側に送り、再生側で両特性を再現する信号を生成して、それらを混合することによって、背景音のある音声入力の場合でも、自然な再生音を出力する。発明のポイントは、音声に特化された符号化モデルの枠組みを大きく切り替えることなく、かつ、限られた情報量（ビットレート）のなかで、うまく雑音の特性を乗せて送るという点で、雑音区間の検出にもそれほどの厳密性は要求されない。

【0012】この発明の復号方法によれば、適応符号帳および固定符号帳の両方または一方からフレーム単位あるいはサブフレーム単位（以降総称してフレーム単位）で取り出した符号ベクトルに、ゲイン符号帳より取り出したゲインを乗算して駆動音源ベクトルを生成し、その駆動音源ベクトルで合成フィルタを駆動して音声信号または音響信号（以降総称して音声信号）を生成する音声の復号方法において、該当フレームが背景雑音区間であるか否かの情報を受け取り、背景雑音区間内のフレームにおいて、合成フィルタの出力信号のパワーを測定し、その長時間平均値を表すパワーレベルを計算し、背景雑



音区間内のフレームにおいて、合成フィルタのフィルタ係数を表すスペクトルパラメータの長時間平均を表す平均スペクトルを計算して、上記平均スペクトルの特性を表すフィルタを、白色雑音で駆動して生成した信号を、上記測定したパワーレベルをもとに振幅調整して、背景雑音の定常成分信号を生成し、上記生成した背景雑音の定常成分信号を、当該フレームが背景雑音区間であるか否かにかかわらず、合成フィルタの出力信号に加算して、再生音声を生成する。

【0013】更に入力信号の特性とは独立した、一定のスペクトル特性、一定の振幅を持つ定常音を、上記平均スペクトルの特性を表すフィルタを白色雑音で駆動して生成した信号とともに、前記背景雑音の定常成分信号として、合成フィルタの出力信号に加算して、再生音声を生成する。また当該フレームが背景雑音区間または子音区間であるか否かの情報を受け取り、駆動音源ベクトルのフレーム毎のパワーを測定し、白色雑音に、上記駆動音源ベクトルのパワーをもとに決定した振幅を乗算して生成した背景雑音の変動成分信号を、当該フレームが背景雑音区間または子音区間である場合に、駆動音源ベクトルに加算する。

【0014】この発明の符号化方法によれば、適応符号帳および固定符号帳の両方または一方からフレーム単位あるいはサブフレーム単位（以降総称してフレーム単位）で取り出した符号ベクトルに、ゲイン符号帳より取り出したゲインを乗算して駆動音源ベクトルを生成し、合成フィルタを駆動して生成した音声信号または音響信号（以降総称して音声信号）と入力音声信号とを比較して、適応符号、固定符号、ゲイン符号を選択する符号化方法において、入力信号を分析して、当該フレームが、背景雑音区間または子音区間に相当するか否かを決定し、当該フレームが、背景雑音区間または子音区間に相当する場合には、上記合成フィルタの出力信号と入力信号の波形歪み最小化に基づく距離尺度と、上記合成フィルタの出力信号と入力信号のパワーレベル差最小化に基づく距離尺度の、加重和を用いるか、またはパワーレベル差最小化に基づく距離尺度のみを用いることによって、ゲイン符号帳を検索して、最適なゲイン符号を選択する。

【0015】なお、この発明の実現にあたっては、信号処理用の専用プロセッサを用いてハードウェア的に実現してもよいし、コンピュータプログラムの形でソフトウェア的に実現してもよい。

【0016】

【発明の実施の形態】以下にこの発明の実施例を、図を用いて説明する。図1は、この発明における符号化方法の機能構成例を示したもので、図3と対応する部分に同一番号を付けてある。また、図2は、この発明における復号方法の機能構成例を示したもので、図4と対応する部分の同一番号を付けてある。この発明のイメージをよ

りわかりやすく説明するため、まず図2の復号方法を先に説明する。またこの発明は音声信号のみならず音楽信号などの音響信号に適用できるが、以下の説明では音声信号で代表して行う。

【0017】この発明による復号方法によって再生される音声信号は、図2において一点鎖線で示される、駆動音源ベクトル生成部35、背景雑音変動成分生成部44、背景雑音定常成分生成部45のそれぞれにおいて生成される信号の全部または一部から構成される。図4に示した従来の復号方法は、上記3つの生成部のうち、駆動音源ベクトル生成部35のみしか有しないため、背景雑音のある音声信号を符号化、復号したときに再生音の自然性が十分でなかった。自然性が十分でない理由は、前述のように、駆動音源ベクトル生成部35のモデルが、人間が音声を生成する機構のモデルに基づいており、背景雑音のある音声のモデル化には必ずしも適当でないことに起因する。これに対して、この発明は、音声の生成モデルによって表現される再生信号（駆動音源ベクトルによる成分）に、雑音モデルによって表現される再生信号が加算されるというモデルに基づいており、加算する雑音のモデルを、短時間で特性が変動する「背景雑音変動成分」、時間的にゆっくりとした周期で特性が変動するか、または変動しない「背景雑音定常成分」の2つの成分で表現する。

【0018】まず、符号化部において送信または蓄積された符号は、符号受信部31において駆動音源符号  $b_c$ 、線形予測パラメータ符号  $b_a$ 、モード情報  $b_m$  に分解される。モード情報  $b_m$  は、雑音付加音声信号において、母音区間、子音区間、無音区間、背景雑音区間といったフレーム毎に区間を分類した情報である。ただし、背景雑音のある入力音声信号を正確に上記区間に分けることは困難であるため、曖昧なフレームは分類不明としてもよい。

【0019】駆動音源ベクトル生成部35は、従来法と同様に、適応符号帳36と固定符号帳37からなり、それぞれの符号帳は、受信した駆動音源符号中の周期符号、固定符号とそれぞれ対応する適応符号ベクトル  $c_a$ 、固定符号ベクトル  $c_r$  を出力する。これらベクトル  $c_a$ 、 $c_r$  は、受信した駆動音源符号中のゲイン符号に対応してゲイン符号帳41から出力されるゲイン  $g_a$ 、 $g_r$  をそれぞれ乗算後、加算されて駆動音源ベクトル  $c$  となる。

【0020】背景雑音変動成分生成部44から出力されるベクトルは、背景雑音の、フレーム単位でパワーやスペクトル特性が変動する成分に対応する。白色雑音生成部51からは、1フレーム長の白色雑音系列ベクトル  $s_1$  が出力される。白色雑音の生成法として一般にはガウス乱数を用いることが多いが、一様乱数を用いたり、あらかじめ乱数系列をテーブルに記憶しておき、フレーム長単位で切り出して用いるといった擬似的な手法

を用いてもよい。

【0021】駆動音源ベクトル生成部35から出力された駆動音源ベクトル  $\mathbf{c}$  は、音源パワー測定部52に入力されて、当該フレームにおける駆動音源ベクトル  $\mathbf{c}$  のパワー  $P_c$  が測定され、測定されたパワー  $P_c$  とモード情報  $\mathbf{b}_m$  に基づいて、白色雑音ベクトル  $\mathbf{s}_1$  と駆動音源ベクトル  $\mathbf{c}$  にそれぞれ乗算される重み  $\mathbf{w}_{s1}$  および  $\mathbf{w}_c$  が、重み作成部53において決定される。白色雑音ベクトル  $\mathbf{s}_1$  と駆動音源ベクトル  $\mathbf{c}$  は、それぞれ重み  $\mathbf{w}_{s1}$  および  $\mathbf{w}_c$  が、乗算部54, 55で乗算された後、加算部56で加算されて、合成フィルタ33への入力ベクトル  $\mathbf{c}_{s1}$  となる。

【0022】なお、重み作成部53において作成される重み  $\mathbf{w}_{s1}$  および  $\mathbf{w}_c$  は、合成フィルタ入力  $\mathbf{c}_{s1}$  のパワーが、駆動音源ベクトル  $\mathbf{c}$  のパワーと同じになるように決める（完全に一致する必要はないが、聴感上同じパワーに聞こえるようにする）。また、当該フレームが母音区間であるか、区間が不明の場合には、 $\mathbf{w}_{s1}$  は0または0に近い小さい値とする。これは、母音区間においては、入力音声は音声の生成モデルに基づく駆動音源ベクトル生成部35で十分良好に表現されるため、背景雑音変動成分は加算する必要がない。また、区間の分類が不確実な場合も、背景雑音変動成分は加算しないほうが無難である。

【0023】図2では背景雑音変動成分生成部44の出力は、合成フィルタ33の手前で駆動音源ベクトルに加算されているが、合成フィルタ33が線形フィルタの場合には、背景雑音変動成分生成部44の出力を合成フィルタに通したものと、駆動音源ベクトルを合成フィルタ33に通したものを加算しても（即ち合成してから加算しても）結果は等価である。

【0024】合成フィルタ33の出力  $\mathbf{y}_{s1}$  は、後処理部34において、従来法と同様に、スペクトル包絡やピッチ成分が強調される。ただし、音声区間以外は強調されると逆に不自然になるため、モード情報  $\mathbf{b}_m$  が母音区間についてのみ従来法と同程度の強調を行い、母音性が低くなるにしたがって、強調の度合いを弱める。合成フィルタ33の出力  $\mathbf{y}_{s1}$  は、背景雑音定常成分生成部45の、平均雑音パワー測定部61へも送られる。平均雑音パワー測定部61では、確実な背景雑音区間

（無音区間を含む）においてのみ、合成フィルタ33の出力信号  $\mathbf{y}_{s1}$  のパワーを測定し、フレーム長に対して十分に長い時間にわたるパワーの平均値を計算する。ここで「確実な背景雑音区間」とは、背景雑音区間でも母音区間や子音区間の疑いのあるフレームは除外することを意味する。この確実な背景雑音区間だけ合成フィルタ33の出力が平均雑音パワー測定部61へ供給されるようにモード情報  $\mathbf{b}_m$  によりスイッチ46が制御される。また「フレーム長に対して十分に長い時間」とは、1秒～数十秒程度がよいと考えられる。長時間平均の計

算の方法としては、バッファにフレーム毎のパワーを記憶しておいて、一定の時間毎に平均をとってもよいし、第  $n$  フレームにおける瞬時パワーを  $P(n)$ 、第  $n$  フレームにおける平均パワーを  $P_{ave}(n)$ 、第  $n-1$  フレームにおける平均パワーを  $P_{ave}(n-1)$ 、 $0 < \alpha < 1$  として、

$$P_{ave}(n) = (1 - \alpha) P_{ave}(n-1) + \alpha P(n)$$

のような逐次更新式を用いて近似的に求めてもよい。なお  $\alpha$  は値が小さいほど長時間平均に相当する。計算された平均雑音パワー  $P_y$  は、重み作成部62に送られ、背景雑音定常成分のパワーを決定する重み  $\mathbf{w}_u$  が計算される。背景雑音定常成分生成部45から出力される信号のパワーは、平均雑音パワー  $P_y$  とほぼ同じになるように決められるが、多少低くなるように決めると、聴感上聞きやすい音になることが多い。なお、合成フィルタ63の出力  $\mathbf{u}_2$  のパワーは平均雑音スペクトル  $\mathbf{a}_u$  に依存するため、重み作成部62で重み  $\mathbf{w}_u$  を求める際には、平均雑音パワー  $P_y$  と合成フィルタ63のフィルタゲインを併用するか、合成フィルタ63の出力  $\mathbf{u}_2$  のパワーを実際に測定してその値をもとに  $\mathbf{w}_u$  を求めるとよい。

【0025】平均雑音スペクトル測定部64では、平均雑音パワー測定部61と同様に、確実な背景雑音区間（無音区間を含む）においてのみ、復号された線形予測パラメータ  $\mathbf{a}_q$  から、フレーム長に対して十分に長い時間にわたるスペクトルの平均値を計算する。このため確実な背景雑音区間のみ線形予測パラメータ  $\mathbf{a}_q$  が平均雑音スペクトル測定部64へ供給されるようにモード情報  $\mathbf{b}_m$  によりスイッチ47が制御される。スペクトルの平均値は、一般に、線形予測パラメータの一種である、線スペクトル対 (LSP) の領域で平均操作を行うことが多いが、ケプストラムやパワースペクトルの領域で平均をとってもよい。平均の計算方法は、上記パワーの平均と同様に、バッファにフレーム毎のスペクトルパラメータを記憶しておいて、一定の時間毎に平均をとってもよいし、逐次更新式を用いてもよい。平均雑音スペクトル測定部64からは、平均雑音スペクトルに対応する、線形フィルタ係数  $\mathbf{a}_u$  が出力され、合成フィルタ63の係数となる。なお、線スペクトル対 (LSP) からの線形フィルタ係数の計算方法は、前述の古井貞熙著「デジタル音声処理」（東海大学出版会）にも記載されている。

【0026】白色雑音生成部65は、白色雑音生成部51と同様に、1フレーム長の白色雑音  $\mathbf{s}_2$  を出力し、出力された白色雑音ベクトル  $\mathbf{s}_2$  は、合成フィルタ63に通されて  $\mathbf{u}_2$  となる。定常音生成部66は、入力信号の性質に依存しない、完全に一定の音  $\mathbf{u}_3$  を出力する。背景雑音定常成分生成部45の出力は、定常音  $\mathbf{u}_3$  の振幅を乗算部67で重み  $\mathbf{w}_{u3}$  を乗算して調整した信号と、合成フィルタ63の出力信号  $\mathbf{u}_2$  とを

加算部68で加算したものに、重み作成部62で作成される重み $w$ を乗算部69で乗算したものととなる。なお、定常音生成部66から出力される信号 $u$ 3は、入力信号とは独立した音であるので、入力信号を符号化し、復号して再生するという立場からいえば、用いても用いなくてもよい。しかし、人間の聴覚特性上、定常な背景雑音は安心感をもたらす、入力信号の特性と必ずしも一致していなくても、より自然に感じることが多い。したがって、定常音生成部66において、ブーンといった低周波の音や、サーといった定常的な白色雑音を生成して、合成フィルタ63の出力レベルに比べて相対的に低いレベルに振幅を調整して加算すると、より自然な背景雑音となる。

【0027】背景雑音定常成分生成部45の出力は、後処理部34の出力信号 $s$ 1eに加算部71で加算されて、再生信号出力となる。ここで注意すべきことは、背景雑音定常成分のレベルは、平均雑音パワーの測定結果 $P_y$ によって決定され、母音区間、子音区間、背景雑音区間といった、フレーム毎のモードにかかわらず、ほぼ一定のレベルで加算されることである。この点は、背景雑音変動成分が、該当フレームのモード情報によって、フレーム毎に、加算される雑音レベルが制御される点と異なっている。

【0028】なお、背景雑音定常成分生成部45の出力は、合成フィルタ33と後処理部34との間で加算してもよいが、後処理部34は音声強調するための処理であるので、音声強調の度合いが大きい場合には、図2に示すように後処理部34の後で加算したほうが処理も簡単で再生音声の自然性も高い。次に、図1を用いて、符号化方法を説明する。

【0029】図2に示したような、復号方法を用いて自然な再生音声を得るためには、符号化側では、従来の符号化方法に加えて、以下の2点を実現しなければならない。1点目は、母音区間、子音区間、無音区間、背景雑音区間といった、フレーム毎のモード分けをして、復号側にモード情報の全部または一部を送ること、2点目は、背景雑音区間において、雑音パワーの情報を復号側に送ることである。

【0030】図1において、入力音声信号 $x$ は、モード判定部81にも送られる。モード判定部81では、入力信号を分析して、区間の特性を表すモード情報 $b$ を出力する。符号化側のモード分けとしては、「母音区間である」「子音区間である」「無音区間である」「背景雑音区間である」の4つに、「よくわからない(不明)」というカテゴリも許すことにする。不明を許すのは、背景雑音が重畳した音声信号を分析した場合、必ず中間的な性質のフレームが存在するため、強制的に4つの区間のどれかに分類してしまうのは自然な音を再生するという立場から適当でないと考えられるからである。ただし、不明を含めた5つのモード情報を受信側に

送ろうとすると、送信情報が無駄になるため、不明モードは符号化側のみで利用して、送信する際には「不明」は母音区間に含めてもよい。上記4つのモード情報を送る場合、一般的には2ビット必要であるが、最初に母音であるかそうでないかを1ビットで表すと、母音以外の区間については、母音区間よりも少ない情報で入力信号を表現できるようになるため、つまり使用ビットに余りが生じるため、この余ったビットで子音/無音/背景雑音の各区間を表すことができ、実質的にはこの発明を実施するためには、従来法と比べてフレームあたり1ビット余分に使用すればよいことになる。フレームあたり1ビットとは、例えばフレーム長が20ミリ秒であれば50ビット/秒に相当し、全体の情報量の4kビット/秒に対し、ごくわずかな情報量増でよい。

【0031】モードを母音区間/子音区間/無音区間/背景雑音区間に分ける手法としては、信号のパワー、パワーの変動分、スペクトル包絡の傾き、ピッチ周期性などを分析して求め、それぞれをしきい値と比較して判断する。また、背景雑音区間の場合は、信号の性質が多岐に渡るため、モードの連続性を考慮したり、過去の背景雑音区間のパワーや性質と比較して、相対的なしきい値を用いるとよい。分析して得られた値とそれぞれのしきい値を比較しても、明確な区間分類ができない場合は、区間が「不明」とする。

【0032】2点目の、背景雑音区間において雑音パワーの情報を復号側に送るために、この発明では、ゲイン符号帳23の探索に、歪み計算部16とパワーレベル差計算部82のそれぞれの出力の加重和を用いて符号帳を検索する。CELP系の符号化方式において、特に背景雑音区間や子音区間では、入力信号のパワーが再生音のパワーに必ずしも反映されない。これは、従来のCELP系符号化方式の符号帳探索が、サンプル単位の波形歪みを小さくすることを念頭において行われていることに対して、符号化のモデルが背景雑音や子音の生成過程に合っていないことに起因する。したがって、従来法を用いた場合、背景雑音区間や子音区間では、再生信号のパワーが、入力信号のパワーを正しく表さないだけでなく、不自然で不安定に変動することが多かった。この発明では、復号側で、合成フィルタ出力 $s$ 1のパワーを計算して、背景雑音レベル $P_y$ を推定するため、従来の符号帳探索方法を用いるのでは、復号側で間違った雑音レベルが推定されてしまうことになる。そこで、この発明における符号化側では、合成フィルタ14の出力 $y$ を従来と同様の波形歪み計算部16に送って、入力信号とのサンプル単位での波形歪み値 $d$ を計算するほか、合成フィルタ出力 $y$ をパワーレベル計算部82にも送り、入力信号のパワーと合成信号 $y$ のパワーの差も計算する。一般にCELP系符号化では、適応符号ベクトル $c_a$ と固定符号ベクトル $c_r$ と、それらに乗算するゲイン $g_a$ 、 $g_r$ の最適な組み合わせを探索

するが、実際には、これらの同時最適値を探索するには膨大な演算量が必要となるため、適応符号帳18と固定符号帳19を先に探索して最適または準最適な適応符号ベクトル  $\mathbf{c}_a$  と固定符号ベクトル  $\mathbf{c}_r$  を決めた後、ゲイン符号帳23を最後に探索することが多い。

【0033】この発明でも上記探索順序によるものとするが、適応符号帳18と固定符号帳19の探索には、従来と同様に、波形歪み  $\mathbf{d}$  の最小化に基づいて符号帳を探索し、ゲイン符号帳23の探索時には、波形歪み  $\mathbf{d}$  とパワーレベル差  $\mathbf{d}_p$  を併用して探索する。重み作成部83では、当該フレームのモード情報  $\mathbf{b}_m$  に基づいて、波形歪み値  $\mathbf{d}$  とパワーレベル差  $\mathbf{d}_p$  のそれぞれの値に乗算部84、85で乗算する重み  $w_d$ 、 $w_p$  を作成する。例えば、母音区間では波形歪み値  $\mathbf{d}$  のみを用いるか波形歪み値  $\mathbf{d}$  を優先する重みの組を使い、子音区間や背景雑音区間ではパワーレベル差  $\mathbf{d}_p$  のみを用いるかパワーレベル差  $\mathbf{d}_p$  を優先する重みの組を使用する。この結果、母音区間では従来と同様の波形歪みの少ない良好な品質の音が、子音区間や背景雑音区間では、合成波形  $\mathbf{y}$  の形状は入力信号  $\mathbf{x}$  の形状と相似性が高く、パワーは入力信号のパワーをできるだけ保存するような音を再生するための駆動音源符号が選択される。なお、「不明」区間は母音区間と子音や背景雑音区間との中間的な性質であるので、ゲイン符号帳23を探索する重みとしてはパワーレベル差  $\mathbf{d}_p$  を重視するような重みを用いるのがよい。乗算部84、85でそれぞれ重み付けされた波形歪み値  $\mathbf{d}$  とパワーレベル差  $\mathbf{d}_p$  は加算部86で加算されて符号帳検索制御部25に入力される。

#### 【0034】

【発明の効果】この発明の効果を調べるために、この発明をコンピュータプログラムによるシミュレーションの形で実現し、実際の音声データを用いて主観品質評価実験を行った。フレーム長は20ミリ秒とし、フレームは2つのサブフレームに分割した。ビットレートは4キロボット/秒で設計した。フレームあたりの詳細なビット配分は、図5に示すように、線形予測パラメータ20ビット、適応符号帳13ビット、固定符号帳34ビットとし、ゲイン符号帳は母音区間では12ビット、母音以外の区間では10ビットとした。モード情報は母音区間で1ビット、母音以外の区間で3ビットである。

【0035】音声データは、修正IRS特性と呼ばれる一般的な電話の特性に準拠したもので、SN比が15d

Bの自動車雑音付加音声と、SN比が30dBのオフィス雑音付加音声を、それぞれ従来法とこの発明を用いて符号化および復号して、それぞれ再生された音を実際に聞いて比較した。被験者は一般人24名で、試験方法は、原音と符号化音声の品質を比較して、非常に悪い(−3)、悪い(−2)、少し悪い(−1)、同品質(0)、少し良い(+1)、良い(+2)、非常に良い(+3)の7段階で評価した。

【0036】図6に試験結果を示す。グラフは低いほど原音に比べて品質が悪いことを示す。図より、この発明を用いた場合の品質は、この発明を用いない従来の方法による品質に比べて大きく改善することが示された。従来法とこの発明による音の性質の差を言葉で表現すると、従来法による再生音は、子音区間と背景雑音区間において、大変に不自然で不快な音であったが、この発明によって再生された音は、自動車騒音とオフィス雑音のそれぞれの雰囲気再現されているうえに、安心して聞ける自然な品質の音であった。また、背景雑音区間だけでなく母音区間においても、音声と背景雑音が混合した、より自然な再生音という観点でこの発明のほうが優れていた。

【0037】なお、この発明を背景雑音のない音声に適用した場合には、「背景雑音」は無音区間として分類され、「背景雑音」の平均レベルは0であると判断されるため、理論的に悪影響を及ぼさないことは言うまでもない。実際の音声を入力した場合も、母音区間と無音区間(背景雑音区間)については従来法と同等の音質が、子音区間については、この発明によるほうが自然な音質であった。

#### 【図面の簡単な説明】

【図1】この発明による符号化装置の機能構成例を示すブロック図。

【図2】この発明による復号化装置の機能構成例を示すブロック図。

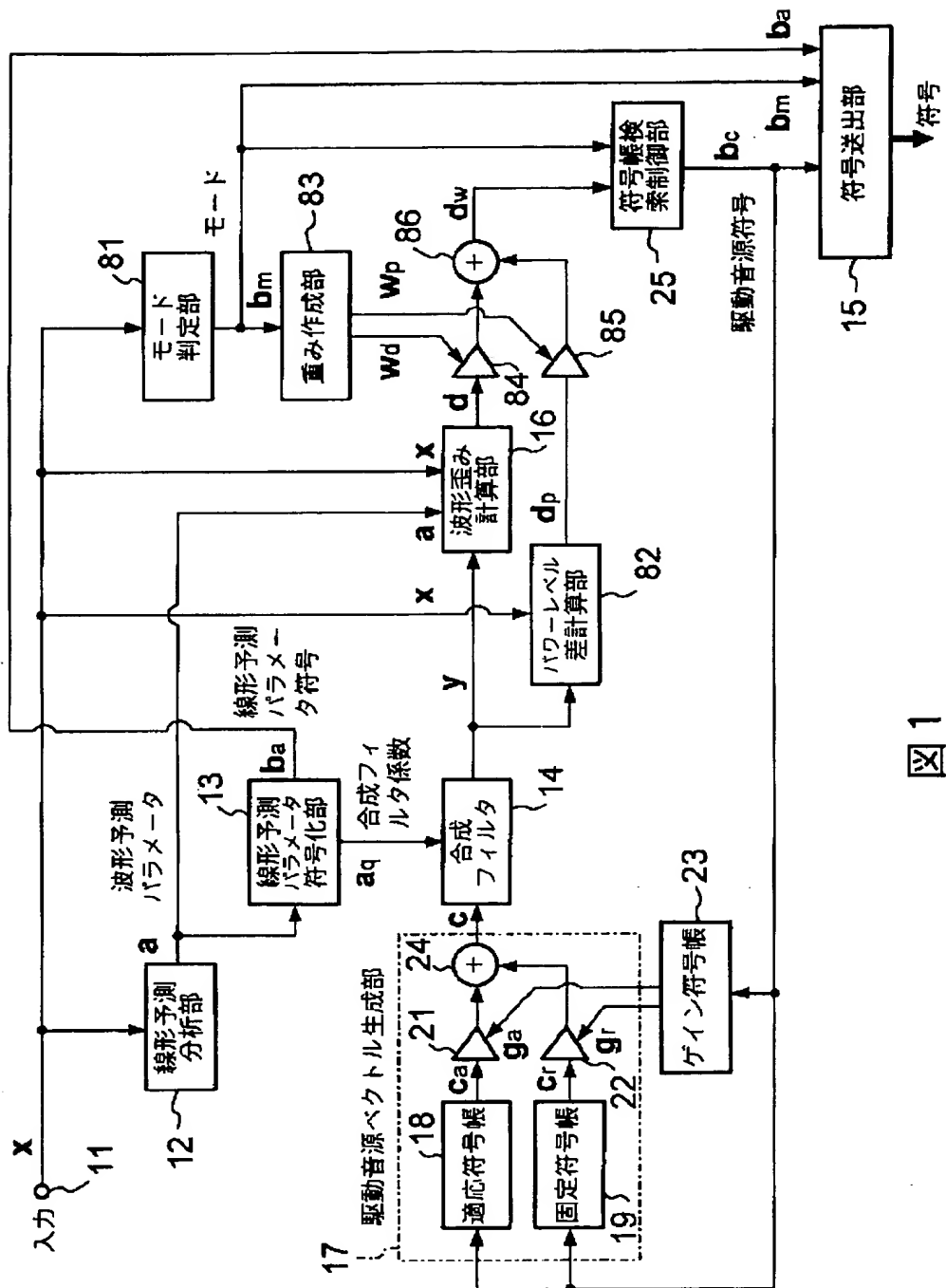
【図3】従来の音声の符号駆動線形予測符号化(Code-Excited Linear Prediction: CELP)装置の機能構成を示すブロック図。

【図4】従来の音声の符号駆動線形予測符号化(Code-Excited Linear Prediction: CELP)に対応する復号装置の機能構成を示すブロック図。

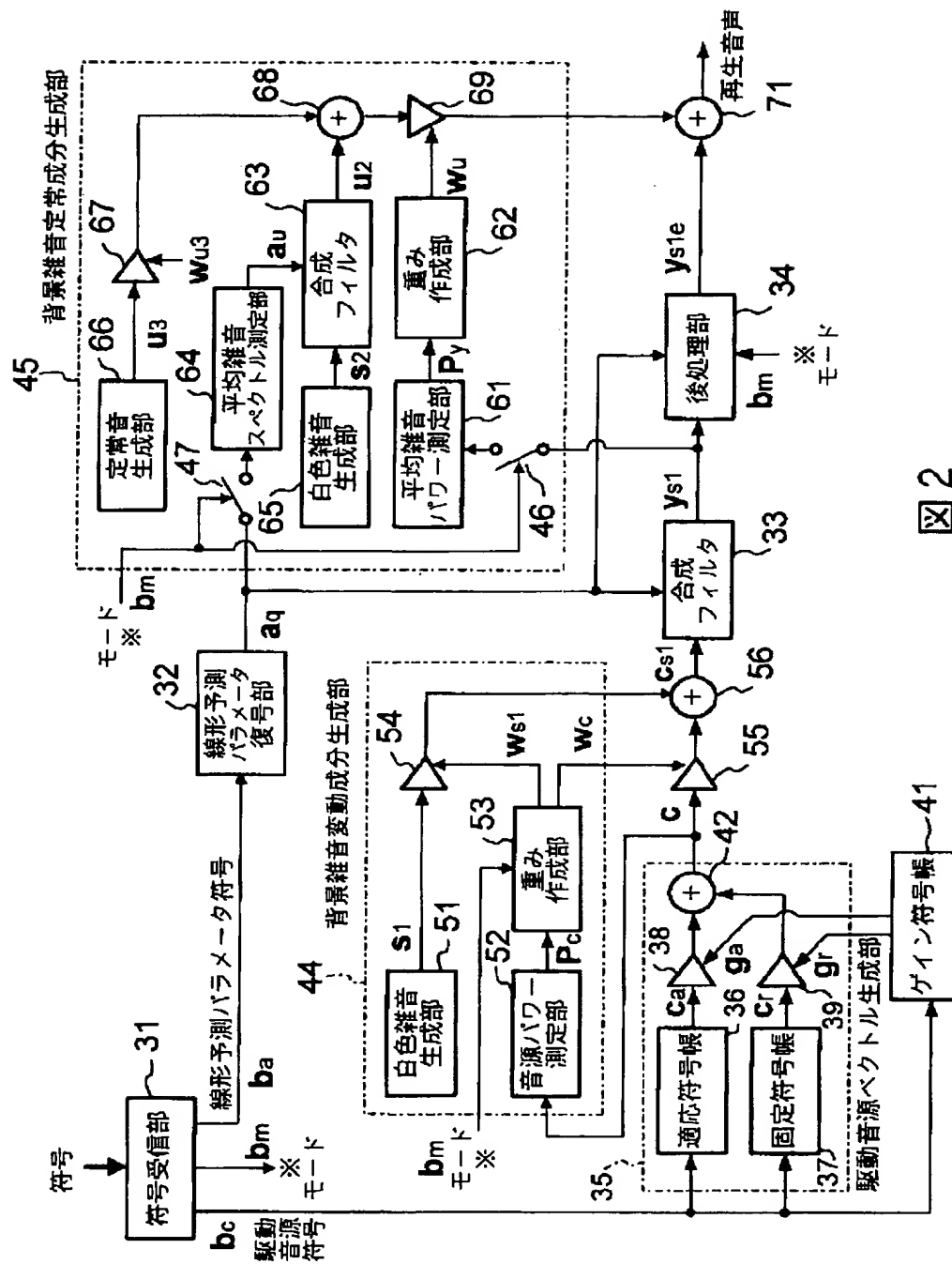
【図5】シミュレーション実験におけるビット配合を示す図。

【図6】シミュレーション実験結果を示す図。

一  
[X]



2  
✗



【図3】

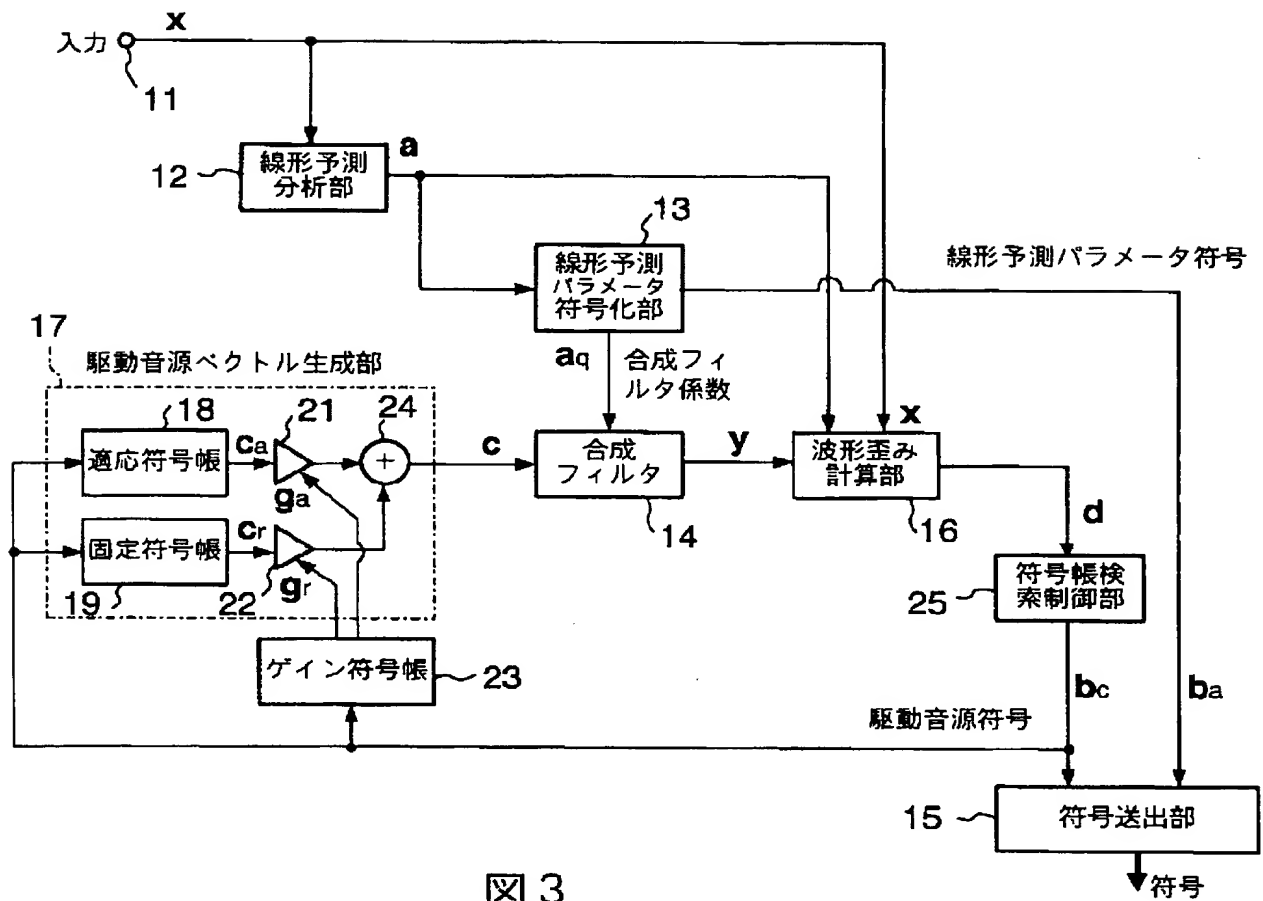


図 3

【図5】

パラメータ	ビット数(ビット/20ms)	
	有声音	有声音以外
線形予測パラメータ	20	
適応符号帳	8+5	
固定符号帳	17+17	
ゲイン符号帳	6+6	5+5
モード情報	1	3
合計	80	

図 5

【図6】

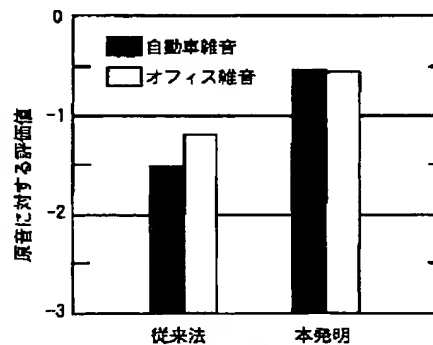


図 6

【図4】

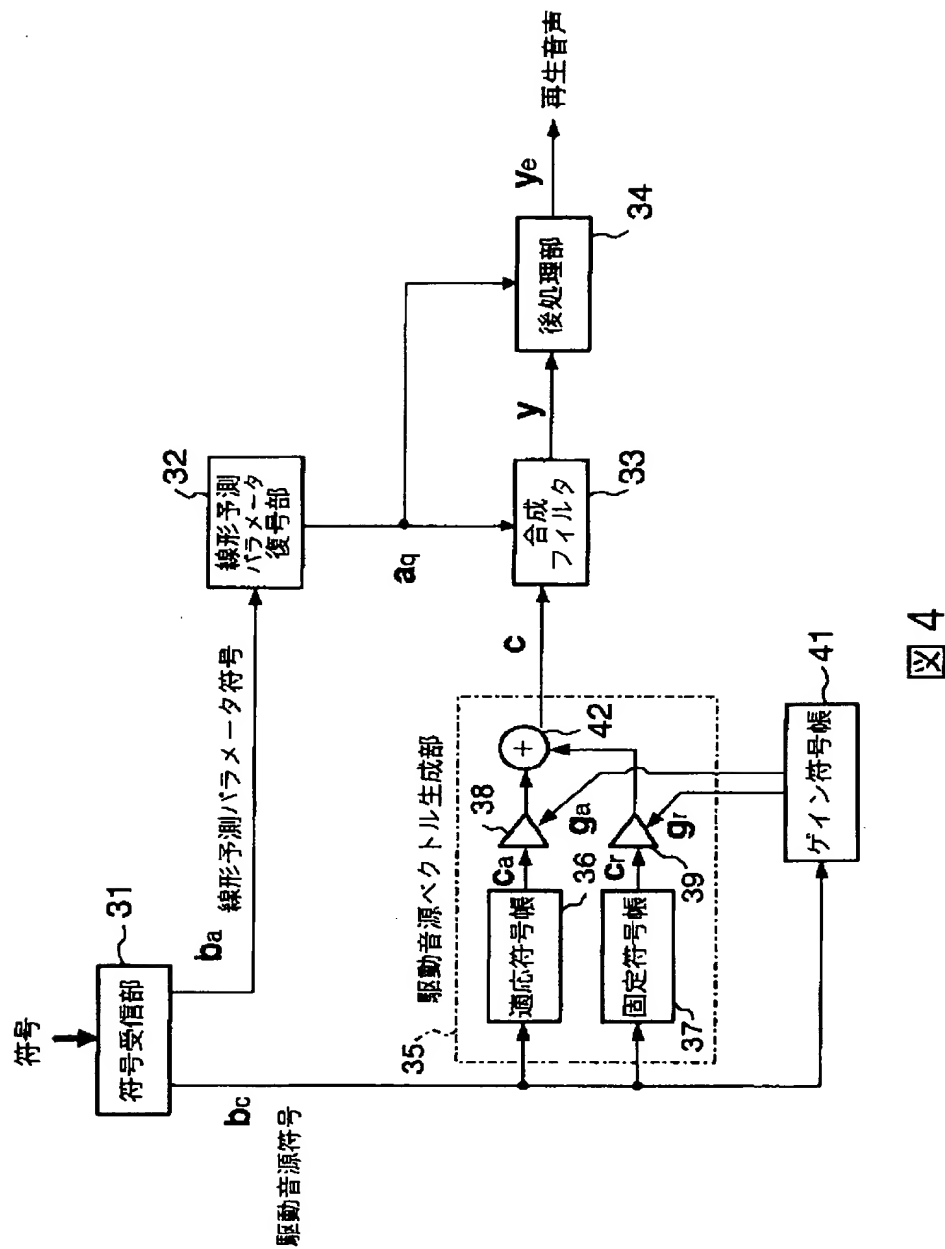


図4

フロントページの続き

Fターム(参考) 5D045 CA04  
 5K041 AA05 AA08 BB08 CC01 CC07  
 DD02 EE00 EE35 EE38 EE51  
 HH27 JJ14  
 9A001 CZ05 EE04 HH15 HH16